

Energy Efficient Cooperative Strategies for Relay-Assisted Downlink Cellular Systems

Part II: Practical Design

Stefano Rini^{*†}, Ernest Kurniawan[†], Levan Ghaghanidze^{*}, and Andrea Goldsmith[†]

^{*} Technische Universität München, Munich, Germany

E-mail: {stefano.rini, levan.ghaghanidze}@tum.de

[†] Stanford University, Stanford, CA, USA

E-mail: ernestkr@stanford.edu, andrea@wsl.stanford.edu

Abstract

In a companion paper [1], we present a general approach to evaluate the impact of cognition in a downlink cellular system in which multiple relays assist the transmission of the base station. This approach is based on a novel theoretical tool which produces transmission schemes involving rate-splitting, superposition coding and interference decoding for a network with any number of relays and receivers. This second part focuses on a practical design example for a network in which a base station transmits to three receivers with the aid of two relay nodes. For this simple network, we explicitly evaluate the impact of relay cognition and precisely characterize the trade offs between the total energy consumption and the rate improvements provided by relay cooperation. These closed-form expressions provide important insights on the role of cognition in larger networks and highlights interesting interference management strategies.

We also present a numerical simulation setup in which we fully automate the derivation of achievable rate region for a general relay-assisted downlink cellular network. Our simulations clearly show the great advantages provided by cooperative strategies at the relays as compared to the uncoordinated scenario under varying channel conditions and target rates. These results are obtained by considering a large number of transmission strategies for different levels of relay cognition and numerically determining one that is the most energy efficient. The limited computational complexity of the numerical evaluations makes this approach suitable for the optimization of transmission strategies for larger networks.

I. INTRODUCTION

Wireless relay nodes hold the promise of drastically increasing both the energy and spectral efficiency of future cellular systems [2]. The demand of transmission rates has increased tremendously between the third and fourth generations of mobile networks due to new functionalities and services provided by handset devices. Rate gains of this order cannot be achieved without increasing bandwidth or the density of the network and relay nodes represent a low-cost way of achieving such increase. Despite the relevance of relays in practical systems, it is not yet clear what system architecture is the most energy efficient and what advantages are provided by cognition in this scenario.

The architecture of relay-assisted downlink cellular system in LTE-A is presented in Fig. 2(a) : the system is comprised of a base station which is interested in communicating to multiple receivers with the aid of the relay nodes. The set of communications between the base station and the relay nodes is termed a *relay link* while the set between relay nodes and receivers is termed an *access link*. We consider the case where no direct link between the base station and receivers exists: this case can be easily obtained by considering an additional relay which is connected to the base station with an infinity capacity channel. In the relay link, transmissions take place over frequency separated channels and are thus non interfering. In the access link, instead, transmissions take place over the same frequency band and therefore are self interfering. When relays cooperate, the access link is analogous to a multi-terminal cognitive channel in which transmitting nodes are able to partially coordinate their transmissions.

The difficulty in designing efficient communication strategies for this system lies in the intrinsic complexity of the communication strategies that can be implemented with even a moderately small number of relays and receivers. Possible ways to circumvent this obstacle is to either consider simple communication protocols for each relay or to assume that relays operate in a uniform manner which does not depend on the network conditions. Two such examples are the so called *relay selection* strategy [3], [4], [5] and a generalization of the amplify-and-forward or decode-and-forward strategies for the relay channel [6].

In [1] we propose a different approach: we generalize the information theoretical derivation of achievable schemes to a network with any number of relays and receivers and obtain the achievable rate region from the analysis of the error probability for this general scenario. This approach is fundamentally different from the previous approaches as it considers all possible ways in which the base station can distribute messages to the relay nodes through cognition and as well as all possible cooperation strategies that can

be implemented at the relay. In particular we focus on the role of partial, or unidirectional, cooperation at the relay which allows for coordinated transmission strategies which have been studied in the context of cognitive channels such as the Cognitive Interference Channel (CIFIC) [7] and the Interference Channel with a Cognitive Relay (IFC-CR) [8].

As a result, the number and the complexity of the scheme considered by this approach has no comparison with other approaches. Moreover, the generation and the evaluation of the schemes has a low computational complexity and it is thus possible to analyze large networks and determine the optimal level of relay cooperation and the most energy efficient transmission strategy.

A. Literature Overview

In the literature, different relay-assisted downlink systems have been proposed and studied. A conceptually simple and yet powerful approach is *relay selection* [3], [4], [5], i.e. selecting the relay which allows for the fastest data transmission between the base station and receivers. Perhaps surprisingly, this scheme provides the same cooperative diversity as more complicated strategies [9]. Space-time coding for relay-assisted transmissions has been considered in [10] and it has been shown to achieve full cooperation diversity at the cost of an increased complexity. The sum rate optimization for amplify-and-forward multi-antenna relay networks is considered in [11], where the “power iterative algorithm” for achieving optimal beamforming is proposed. Achievable rate regions in two-way relay communications is discussed in [12]. Combined strategies using amplify-and-forward, decode-and-forward, compress-and-forward with superposition coding and other encoding and decoding techniques are compared for different scenarios.

In the context of energy efficiency, virtual MIMO systems have been considered to exploit the broadcast nature of the medium and distribute messages from the base station to the relay nodes. In this case it is possible to analyze the performance of a system where relay nodes opportunistically decode the transmissions that can be overheard over the network. When multiple relays are able to decode multiple transmissions, they are then able to act as a virtual MIMO system in the next hop. This can be incorporated into multi-hop transmission schemes to study the energy efficiency as in [13] and [14]. Energy efficiency for relay-assisted LTE-A networks is studied in [15], [16] where it is shown that a multi-cast cooperative scheme in relay networks results in more efficient energy consumption compared to a conventional two-hop approach. This approach does not consider the cooperation among multiple relays and considers only limited transmission strategies, but it is a first approach to this problem. Another approach to energy efficiency analysis of cooperative relaying is considered in [17]-[18]: here an opportunistic scheme is

proposed, which relies on the channel state information (CSI) knowledge to decide which subset of the relay nodes cooperate.

B. Contributions

Previous literature focuses on the case of either no relay cooperation or full relay cooperation. We focus here the intermediate case of partial and unidirectional transmitter cooperation. In information theory, these forms of cooperation are referred to as cognition as in the CIFC [7] and the IFC-CR [19]. The two channels idealize the ability of some transmitter to learn the message of the other user thanks to the broadcast nature of the wireless medium.

In a companion paper [1], we propose a new theoretical framework to design optimal transmission strategies for the case where each relay has knowledge of multiple messages. We are able to derive the achievable rate region of any scheme transmission scheme at the relay involving rate-splitting, superposition coding and interference decoding. The schemes to be considered on the access link depend on the message decoded at the relay nodes, which we refer to as *level of cognition*.

Although very powerful, this approach is too general to provide much insight on the important features of energy efficient transmission schemes. In the following we specialize the approach of [1] to the network where a base station communicates to three receivers with the aid of two relay nodes. For this specific network we provide closed-form expression for the achievable rate regions which provide significant insights on the trade off between cooperation and interference management at the relays. We also present a simulation setup which shows how the theoretical results in [1] can be efficiently implemented. This shows that the proposed approach can indeed be used to analyze larger networks.

The contributions in the paper can be summarized as follow:

Explicit Characterization for Power Consumptions: Using the results in [1], we derive explicit characterizations for the power consumption of several transmission strategies and provide insights on the tradeoff between power consumption at the base station and cooperation at the relays. By comparing the performance of different schemes, we show that cooperation is not only useful in increasing the power efficiency relative to the uncoordinated case but indeed is necessary to attain high data rates.

Numerically Determine Optimal Strategies We automate the derivation of achievable rate regions in [1] and through computer simulation we determine the most energy efficient transmission strategies for a number of channel realizations. In each case, we consider a number of possible achievable strategies

in the order of the thousands for the channel model under consideration. The minimization of the power consumption necessary to achieve a target rate for each scheme can be performed with almost linear complexity. Indeed the fact that our approach is able to consider a large number of transmission schemes and optimize each scheme with reasonable complexity makes it very suitable for larger networks.

C. Paper Organization

Section II introduces the network model under consideration: the relay-assisted downlink cellular system in with two relays and three receivers. The problem of energy minimization this network is introduced in Sec. III. Lower bounds to the energy efficiency are provided in Sec. IV. Explicit characterizations for the network topology under consideration are provided in Section V. The results of the numerical optimization are provided in Section VI. Finally, Section VII concludes the paper.

D. Notation

In the remainder of the paper we adopt the following notation:

- variables related to the Base Station (BS) are indicated with the superscript BS, moreover i is the index related to BS,
- variables related to the Relay Nodes (RN) are indicated with the superscript RN, moreover j is the index related to RNs,
- variables related to the Receivers (RX) are indicated with the superscript RX, moreover z is the index related to RXs,
- $\mathcal{C}(\Sigma) = 1/2 \log(|\Sigma \Sigma^H + \mathbf{I}|)$ where X is a vector of length k of jointly Gaussian random variables and $|A|$ indicates the determinant of A ,
- A_{ij} element of the matrix A in row i and column j ,

II. CHANNEL MODEL

In the following we consider a relay-assisted downlink cellular system with two relays and three receivers. In this model, also depicted in Fig. 1, the Base Station (BS) is interested in communicating to three receivers (RXs) with the aid of two Relay Nodes (RNs). This model is a special case of the general model in [1] in which $N_{\text{RN}} = 2$ and $N_{\text{RX}} = 3$, in particular each RX $z \in \{1, 2, 3\}$ is interested in the message W_z at rate R_z which is known at the BS and is to be transmitted reliably and efficiently to RXs

through the RNs 1 and 2. The channel inputs at the Base Station (BS) are X_1^{BS} and X_2^{BS} each subject to the power constraint is $E[|X_1^{\text{BS}}|] + E[|X_2^{\text{BS}}|] \leq P^{\text{BS}}$. The channel outputs at Relay Node(RN) one and two

$$Y_1^{\text{RN}} = d_{11}X_1^{\text{BS}} + Z_1^{\text{RN}} \quad (1a)$$

$$Y_2^{\text{RN}} = d_{22}X_2^{\text{BS}} + Z_2^{\text{RN}}. \quad (1b)$$

In a similar fashion, the channel input at the RNs are X_1^{RN} and X_2^{RN} subject to the power constraints $E[|X_1^{\text{RN}}|] \leq P_1^{\text{RN}}$ and $E[|X_2^{\text{RN}}|] \leq P_2^{\text{RN}}$. The channel outputs at the RXs are

$$Y_1^{\text{RX}} = h_{11}X_1^{\text{RN}} + h_{12}X_2^{\text{RN}} + Z_1^{\text{RX}} \quad (2a)$$

$$Y_2^{\text{RX}} = h_{21}X_1^{\text{RN}} + h_{22}X_2^{\text{RN}} + Z_2^{\text{RX}} \quad (2b)$$

$$Y_3^{\text{RX}} = h_{31}X_1^{\text{RN}} + h_{32}X_2^{\text{RN}} + Z_3^{\text{RX}}. \quad (2c)$$

Each noise term Z has variance one and the channel coefficient can take any complex value. A graphical representation of the channel is provided in Fig. 1.

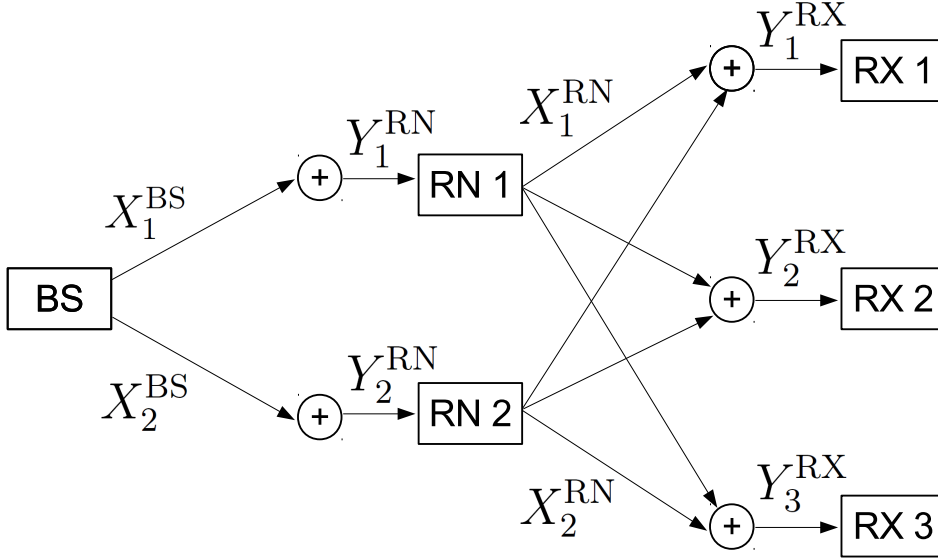


Fig. 1. The relay-assisted downlink cellular system in with two relays and three receivers.

The transmission between the BS and the RNs as well as the transmission between the RNs and the RXs takes place over N channel transmissions. Each message W_z is uniformly distributed in the interval $[1 \dots 2^{N R_z}]$. Let W indicate the vector containing all the messages to be transmitted, i.e. $W = [W_1 \ W_2 \ W_3]$ and R the vector containing the rate of each message, i.e. $R = [R_1 \ R_2 \ R_3]$. Additionally let W_j^{RN} be the

set of messages decoded at relay node $j \in \{1, 2\}$ and define $W^{\text{RN}} = [W_1^{\text{RN}} \ W_2^{\text{RN}}]$. A transmission on the relay link is successful if there exists an encoding function at the BS and a decoding function at each RN such that each relay can successfully decode the message in W_j^{RN} with high probability. Similarly, a transmission on the access link is successful if there exists an encoding function at each RN and a decoding function at each RX such that each receiver $z \in \{1, 2, 3\}$ can decode the message W_z reliably. More formally, let $\widehat{W}_z^{\text{RN}_j}$ be the estimate of W_z at relay j and \widehat{W}_z the estimate of W_z at receiver z over N channel transmissions, then a communication error occurs when there exist $\widehat{W}_z^{\text{RN}_j} \neq W_z$ or $\widehat{W}_z \neq W_z$ for some noise realization over the relay link or the access link.

A rate vector R is said to be achievable if, for any $\epsilon > 0$, there is an N such that

$$\max_z \max_{W_j^{\text{RN}}} \mathbb{P} \left[\widehat{W}_z^{\text{RN}_j} \neq \widehat{W}_z \neq W_z, \right] \leq \epsilon.$$

Capacity is the closure of the union of the sets of achievable rates.

In the following we consider the problem of minimizing E_{TOT} , the total energy required to achieve a rate vector $R = [R_1 \ R_2 \ R_3]$ defined as

$$E_{\text{TOT}} = \frac{P_{\text{TOT}}}{R_1 + R_2 + R_3}, \quad P_{\text{TOT}} = P^{\text{BS}} + \sum_{j=1}^3 \mu_j \frac{P_j^{\text{RN}}}{R_j}. \quad (3a)$$

III. PROBLEM DESCRIPTION

In this section we provide some initial insight on the role of cooperation for the network model in Sec. II. We are interested in characterizing the energy advantages provided by cooperation and cognition when employing superposition coding, interference decoding and rate-splitting. When the base station distributes the same message to multiple RNs, the access link becomes cognitive channel in the sense of the Cognitive InterFERENCE Channel (CIFC) and the InterFERENCE Channel with a Cognitive Relay (CIFC-CR), that is, transmitting nodes have partial or unidirectional knowledge of each others' messages and can thus perform partially coordinated transmissions. Transmitter cooperation has been shown to provide substantial improvement on the rate performance [20] as well as suggesting new and surprising cooperative transmission schemes [21]. In order to obtain the performances of cognitive channels though, it is necessary to distribute the same message to multiple RNs which consumes more power in the relay link. Indeed we are interested in characterizing the tension between increase of energy in the relay link and the benefits provided by cognition in the access link. In general, there exists an optimal level of cognition at the relay nodes for which the additional energy consumption at the relay link allows for the most beneficial

performance improvements in the access link. Our aim is to quantify the amount of RN cooperation that is needed to attain the largest energy efficiency and determine what coding choices attain it.

Let us consider the scenario in which the BS distributes each message to only one RN, as illustrated in Fig. 2.A. In this case, the power utilization at the BS is minimal but no cooperation between the relays is possible as no message is known at both. The lack of cooperation among RNs corresponds, in general, to a higher power consumption on the access link as there is no diversity gain at the receivers: when multiple RNs send the same signal, we obtain a combining gain at the receiver. Cooperation provides some additional advantages beyond combining: by knowing the same message, RNs can each superimpose the codewords of other users over the codeword of the common message. This, in general, achieves larger energy savings than sending each codeword separately. Consider now the message allocation at the RNs in Fig. 2.B: since message W_1 and W_2 are distributed to both RNs, it is possible for the RNs to cooperate in transmitting these messages and the codeword to transmit W_3 can be superimposed over the common codewords. The transmission of a message to multiple RNs, however, utilizes more power on the relay link and the advantages attained on the access link must be weighted against this additional power consumption.

The aim of the paper is to investigate the tradeoff between cooperation and power consumption when considering classical information theoretical achievable strategies. We attempt to identify the message allocation at the relay nodes and the relay cooperation strategy which allows for the lowest overall power consumption.

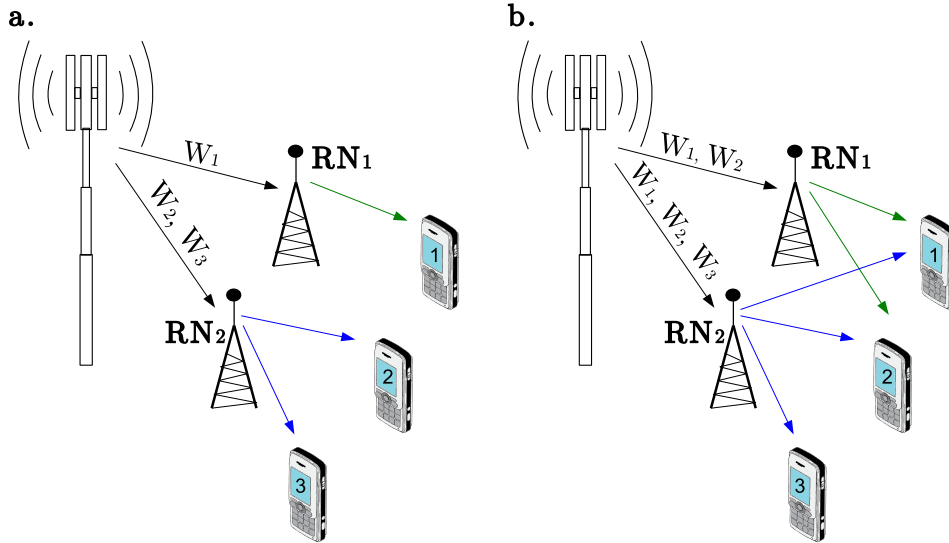


Fig. 2. An example comparison of using cooperative transmission strategy and transmitting without cooperation.

In the network of Sec. II, the possible levels of relay cognition are:

- **No cooperation: each message is known only at one RN.** The RNs code independently and the RN that knows both message can apply superposition coding among these codewords.

This scenario minimizes the power consumption at the relay link but no gain from cooperation is possible in the access link.

- **Partial Cooperation I: a message is known at both RN.**

The codeword of the common message acts as a base codewords on top of which the other two codewords can be superimposed. When superposition coding is applied, the RXs decoding the top codewords must also decode the bottom one: this imposes additional rate constraint on the bottom codeword but improves the energy efficiency in sending the top codewords.

- **Partial Cooperation II: two messages are known at both RN.**

The codeword of the two common messages can be superimposed one on top of the other and act as base codewords for the message known only at one RN.

This improves on the energy gains offered by superposition coding with respect to the previous scenario but consumes additional power on the relay link.

- **Full Cooperation: all messages are known at all RNs.** In this case the power consumption at the base station is maximal but the RNs can cooperate on the transmission of every codeword.

The different transmission strategies corresponding to the different level of cooperation are depicted in

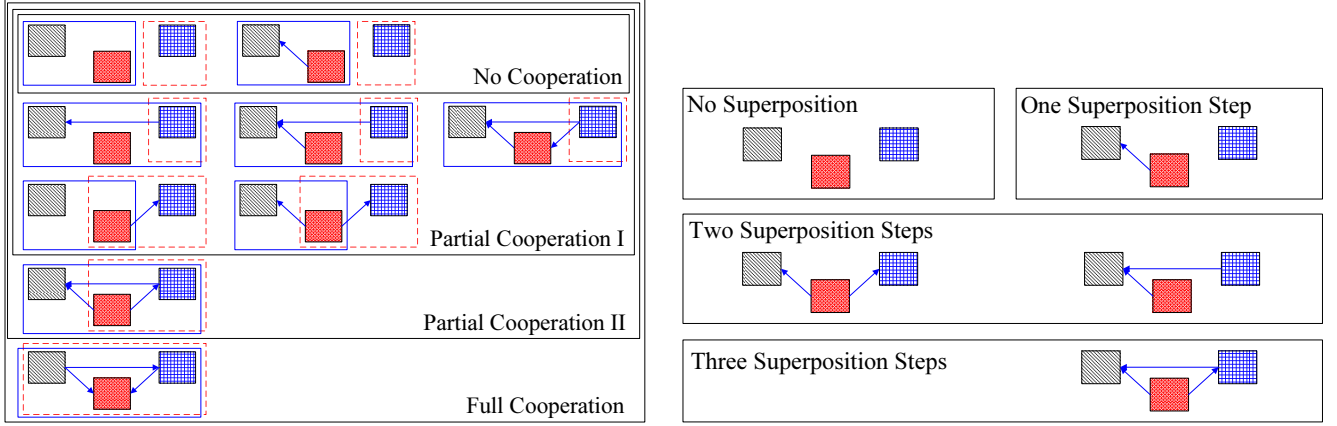
Fig. 3(a). Each hatched box represents the codeword of a message: codewords are not labeled to account for the symmetry of the possible W_j^{RN} . The boxes encircled by the solid blue line are known at one RN while the boxes encircled by a dotted red line by the other RN (no labeling is used to account for symmetry).

Since the different cases are listed with increasing number of shared messages, at each step the transmission strategies of the previous cases are also viable.

Arrows connecting boxes represent the superposition coding: the orientation is from the bottom codewords toward the top one.

The decoding of the top codeword requires also the decoding of the bottom codewords. For this reason, superposition coding restricts the rate of the base codeword to the highest rate that can be reliably decoded by all the receivers which must decode this codeword. Consequently applying superposition coding does not always result in a rate advantage: this is the case only when the decoding of the base codeword at a non-intended receiver does not limit the achievable rates at the intended one.

To distinguish among possible application of superposition coding, we can classify the possible encoding strategies depending on the number of superposition coding steps they involve: one, two or three. The possible encoding strategies in each case are depicted on Fig. 3(b).



(a) Different superposition coding schemes for different levels of cooperation. (b) Different superposition coding schemes with no, one, two or three superposition coding steps.

Fig. 3. Cognition and cooperation for the relay-assisted downlink cellular system in with two relays and three receivers.

Different schemes require at least a certain level of cooperation among relays and the same scheme can be implemented for different levels of cooperation. This classification uniquely identifies the number of interfering codeword that a receiver is required to decode.

IV. LOWER BOUNDS TO THE ENERGY EFFICIENCY

In this section we specialize the results in [1, Th. V.1.], [1, Th. V. 2.] and [1, Lem. V. 3.] to the channel model in Sec. II. Although simple, this example provides the reader with the basic intuition on the structure and role of these bounds.

Lemma IV.1. Relay Link Capacity for the Two Relays and Three Receivers Case *For a given message allocation W^{RN} the capacity of the relay link for the channel model in Sec. II is*

$$\sum_{z \in W_1^{\text{RN}}} R_z \leq \mathcal{C}(|d_{11}|^2 P_1^{\text{BS}}) \quad (4a)$$

$$\sum_{z \in W_2^{\text{RN}}} R_z \leq \mathcal{C}(|d_{22}|^2 P_2^{\text{BS}}), \quad (4b)$$

for any P_1^{BS} and P_2^{BS} such that $P_1^{\text{BS}} + P_2^{\text{BS}} = P^{\text{BS}}$.

Now for the access link outer bounds we have

Lemma IV.2. Access Link Outer Bound for the Two Relays and Three Receivers Case

Given a message allocation W^{RN} , an outer bound to the achievable rates on the access link for the channel in Sec. II is

$$R_1 \leq \mathcal{C}(H_{11}A_{11} + H_{12}A_{21}) \quad (5a)$$

$$R_2 \leq \mathcal{C}(H_{21}A_{12} + H_{22}A_{22}) \quad (5b)$$

$$R_3 \leq \mathcal{C}(H_{31}A_{13} + H_{32}A_{23}) \quad (5c)$$

$$R_1 + R_2 \leq \mathcal{C}\left(\begin{bmatrix} H_{11} & H_{12} \\ H_{21} & H_{22} \end{bmatrix} \begin{bmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{bmatrix}\right) \quad (5d)$$

$$R_1 + R_3 \leq \mathcal{C}\left(\begin{bmatrix} H_{11} & H_{12} \\ H_{31} & H_{32} \end{bmatrix} \begin{bmatrix} A_{11} & A_{13} \\ A_{21} & A_{23} \end{bmatrix}\right) \quad (5e)$$

$$R_2 + R_3 \leq \mathcal{C}\left(\begin{bmatrix} H_{21} & H_{22} \\ H_{31} & H_{32} \end{bmatrix} \begin{bmatrix} A_{12} & A_{13} \\ A_{22} & A_{23} \end{bmatrix}\right) \quad (5f)$$

$$R_1 + R_2 + R_3 \leq \mathcal{C}(HA), \quad (5g)$$

union over all the possible matrices A as defined in [1, eq. (27)].

From Lem. IV.1 and Lem. IV.2 we obtain a lower bound on the energy consumption by using the fact that determining capacity is the dual problem of determining power efficiency.

Lemma IV.3. Energy Consumption Lower Bound

A lower bound to energy required for achieving the rate vector R in the channel model in Sec. II is obtained as the minimum over W^{RN} defined as

$$W_1^{\text{RN}}, W_2^{\text{RN}} \subseteq \{1, 2, 3\} \quad (6a)$$

$$W^{\text{RN}} = \{W_1^{\text{RN}}, W_2^{\text{RN}}\}, \quad (6b)$$

of P^{BS} and $[P_1^{\text{RN}} \dots P_{N_{\text{RN}}}^{\text{RN}}]$ such that R is on the boundary of (4) and (5) respectively.

V. ACHIEVABLE SCHEMES

In this section we provide some closed-form expressions of achievable rate region which provide concrete examples of the tradeoffs between cooperation and power consumption that emerge in the set up we consider. This section also sets the stage for the numerical simulations in Sec. VI where some of the schemes presented here are simulated numerically.

In particular we consider four schemes

- **Non-cooperative scheme with two active relays:** this scheme minimizes the power consumption on the relay link by distributing the one message to one relay and two messages to the other.
- **Non-cooperative scheme with two active relays and interference decoding:** even when non cooperative schemes are implemented, interference decoding can be used to boost the energy efficiency
- **Partially cooperative scheme:** in which the relays cooperate in transmitting at least one message
- **Partially cooperative scheme with interference decoding:** which combines both cooperative gains and interference decoding gains.

The schemes above are taken without rate-splitting, so that Γ is a binary matrix. In the following, to avoid having to specify the rate-splitting matrix in each different scenario, we adopt the notation $U_{i \rightarrow j}(W_z)$ to indicate that the message W_z is embedded in the codeword $U_{i \rightarrow j}^N$.

We again focus on the simple channel model in Sec. II and additionally focus on the symmetric case in which

$$H = \begin{bmatrix} 1 & b \\ a & a \\ b & 1 \end{bmatrix}, \quad (7)$$

for $a, b \in \mathbb{R}^+$ and

$$d_{11} = d_{22} = 1, \quad (8)$$

With this specific choice of channel parameters, we have that

- by fixing $a = 0$, the access channel reduces to an interference channel,
- by fixing $b = 0$, no communication is possible between RN 1 and RX 3 and between RN 2 and RX 1,
- when one of the relay has knowledge of all messages and the other none, the channel reduces to a degraded Gaussian Broadcast Channel (BC),
- when $a = 0$ and RNs know independent messages, the access channel reduces to an interference channel,
- when $a = 0$ and a relay node knows the message of the other, the access channel reduces to a cognitive interference channel,

A graphical representation of the channel under consideration is provided in Fig. 4.

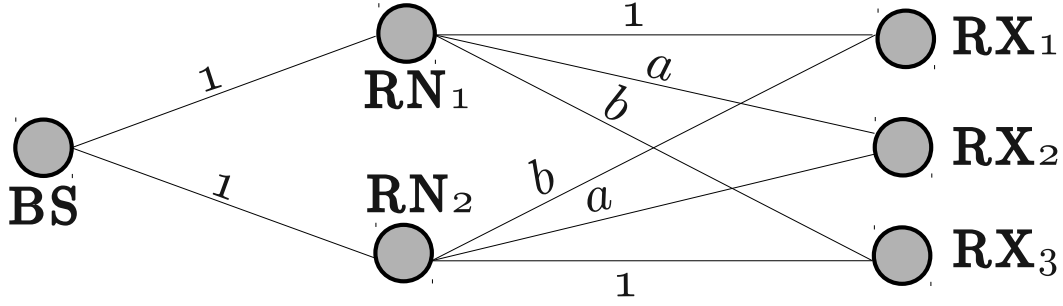


Fig. 4. The symmetric channel model considered in Sec. V.

We also focus on the symmetric rate problem, that is we fix $R_{\text{sym}} = R_1 = R_2 = R_3$. By choosing a highly symmetric model and symmetric rates, we make it possible to derive simple explicit characterizations which lend themselves to intuitive interpretations and a very simple graphical representation.

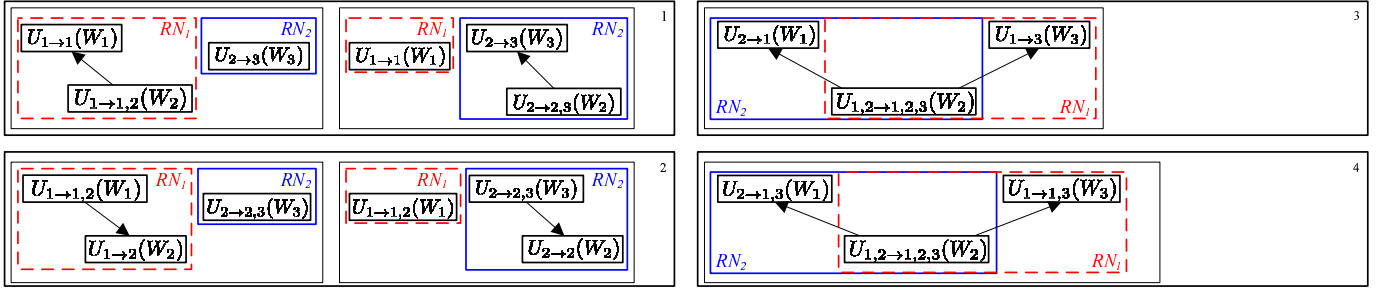


Fig. 5. The four CGRAS presented in Sec. V.

A. Non-cooperative scheme with two active relays

The power consumption at the BS is minimized when transmitting one message to a RN and two messages to the other RN, in which case $P^{\text{BS}} = 16^{R_{\text{sym}}} + 4^{R_{\text{sym}}} - 2$. In this case, superposition coding can be applied at the relay node that knows two messages: the order of superposition of the codeword is chosen according to the SINR toward the two RXs for which the messages are intended. This scheme is in general efficient also on the access link but it is interference limited at the decoders which treat the interference as noise. For this reason, it may not be possible to find a power allocation which achieves a given target rate.

Let's consider the case in which RN 1 has knowledge of W_1 and RN 2 has knowledge of W_2 and W_3 (other message combinations are easily obtained in a similar manner and also considering the symmetry of the channel). This assignment is advantageous in the case $a < 1$ and $b < 1$ in which case it is more

convenient to send W_1 through RN 1 instead of RN 2 and W_3 through RN 2. Also, since $a < 1$, RN 2 it is advantageous to superimpose the codeword for W_3 over the one for W_2 .

This corresponds to the CGRAS number 1 in Fig. 5 (same as CGRAS number 3 in Fig. 7(a)). With this coding choice the following rate region is achievable

$$R_1 \leq \mathcal{C} \left(\frac{P_{11}}{1 + b^2(P_{22} + P_{23})} \right) \quad (9a)$$

$$R_2 \leq \mathcal{C} \left(\frac{a^2 P_{22}}{1 + a^2 P_{11} + a^2 P_{23}} \right) \quad (9b)$$

$$R_2 + R_3 \leq \mathcal{C} \left(\frac{P_{22} + P_{23}}{1 + b^2 P_{11}} \right) \quad (9c)$$

$$R_3 \leq \mathcal{C} \left(\frac{P_{23}}{1 + b^2 P_{11}} \right), \quad (9d)$$

with the assignment

$$X_1 = \sqrt{P_{11}} U_{1 \rightarrow 1}(W_1) \quad (10a)$$

$$X_2 = \sqrt{P_{22}} U_{2 \rightarrow 23}(W_2) + \sqrt{P_{23}} U_{2 \rightarrow 3}(W_3), \quad (10b)$$

which has a total power consumption at the access link of

$$P_1^{\text{RN}} + P_2^{\text{RN}} = P_{11} + P_{22} + P_{23}. \quad (11)$$

When $a \geq 1 \geq b$, the sum rate bound (9c) is never redundant which means that the region in (9) has two Pareto optimal corner points. With the assumption, $R_{\text{sym}} = R_1 = R_2 = R_3$, the required transmit power can then be calculated as:

$$P_{11} \geq \mathcal{C}^{-1}(R_{\text{sym}}) (1 + b^2(P_{22} + P_{23})) \quad (12a)$$

$$P_{22} \geq \mathcal{C}^{-1}(R_{\text{sym}}) (a^{-2} + P_{11} + P_{23}) \quad (12b)$$

$$P_{22} + P_{23} \geq \mathcal{C}^{-1}(2R_{\text{sym}}) (1 + b^2 P_{11}) \quad (12c)$$

$$P_{23} \geq \mathcal{C}^{-1}(R_{\text{sym}}) (1 + b^2 P_{11}), \quad (12d)$$

for $\mathcal{C}^{-1}(x) = 4^x - 1$.

When combining equations (12a) and (12c) we obtain

$$P_{11} (1 - b^4 \mathcal{C}^{-1}(2R_{\text{sym}}) \mathcal{C}^{-1}(R_{\text{sym}})) \geq \mathcal{C}^{-1}(R_{\text{sym}}) (1 + b^2 \mathcal{C}^{-1}(2R_{\text{sym}})). \quad (13)$$

A necessary condition for P_{11} to be positive is for parameter b to satisfy

$$b^4 \leq \frac{1}{\mathcal{C}^{-1}(2R_{\text{sym}})\mathcal{C}^{-1}(R_{\text{sym}})}, \quad (14)$$

similarly, by combining (12a), (12b) and (12d) we obtain that a necessary condition for P_{11} to be positive is

$$b^2 \leq \frac{1}{\mathcal{C}^{-1}(R_{\text{sym}})(1 + \mathcal{C}^{-1}(R_{\text{sym}}))}. \quad (15)$$

It can be verified that P_{22} and P_{23} can always be determined when P_{11} is feasible. This consideration is not surprising: since RX 1 and RX 3 treat the interference as noise, large rates can be achieved only when the level of the interference, which is controlled by b in both cases, is small. For b to be small it is necessary for R_{sym} to be small as prescribed by (14). Other cases of parameter values can be similarly analyzed with similar conclusions: this scheme employs no cooperation among the relay and it is thus power efficient on the relay link but does not achieve arbitrarily large rates as it is interference limited.

B. Non-cooperative scheme with two active relays and interference decoding

The achievable scheme in Sec. V-A can be improved upon by allowing some receivers to decode an interfering codeword. Consider again (9b): in this case the rate R_2 is limited by the power of the codeword for W_1 : a possible solution for this problem is to allow interference decoding at RX 2.

To show the advantages provided by interference decoding consider the case where $a > 1 > b$: in this case a well performing achievable scheme that does not consider cooperation is the scheme where RN 2 transmits W_1 and RN 1 transmits W_3 . W_2 can be sent to RN 1 and, since $a > b$, superimposed over the codeword for W_3 (in this, the scheme considered here differs from the scheme in Sec. V-A, in which W_3 is superimposed over W_2). Since $a > b$, the transmission from RN 2 toward RX 1 interfere with the transmission from RN 1 to RX 2: in this case it is convenient to have RX 2 decode W_1 . With this choice, while RX 1 and RX 3 only decode their intended message, RX 2 decodes all the messages.

This corresponds to the CGRAS number 2 in 5 (same as the CGRAS number 3 in Fig. 7(a)) with

achievable rate region:

$$R_1 \leq \mathcal{C} \left(\frac{P_{11}}{1 + b^2(P_{22} + P_{23})} \right) \quad (16a)$$

$$R_1 + R_2 + R_3 \leq \mathcal{C} (a^2(P_{11} + P_{22} + P_{23})) \quad (16b)$$

$$R_2 + R_3 \leq \mathcal{C} (a^2(P_{22} + P_{23})) \quad (16c)$$

$$R_1 + R_2 \leq \mathcal{C} (a^2(P_{11} + P_{22})) \quad (16d)$$

$$R_2 \leq \mathcal{C} (a^2 P_{22}) \quad (16e)$$

$$R_1 \leq \mathcal{C} (a^2 P_{11}) \quad (16f)$$

$$R_3 \leq \mathcal{C} \left(\frac{P_{23}}{1 + b^2 P_{11} + P_{22}} \right), \quad (16g)$$

with the assignment

$$X_1 = \sqrt{P_{11}} U_{1 \rightarrow 12}(W_1) \quad (17a)$$

$$X_2 = \sqrt{P_{22}} U_{2 \rightarrow 2}(W_2) + \sqrt{P_{23}} U_{2 \rightarrow 23}(W_3), \quad (17b)$$

which has a total power consumption at the access link of

$$P_1^{\text{RN}} + P_2^{\text{RN}} = P_{11} + P_{21} + P_{23}. \quad (18)$$

In this case the rate bounds (16a) and (16g) are still interference limited as in the scheme in (9) but the remaining bounds are not. On the other hand though, the additional bounds of (16b), (16c) and (16d) are not present in (9). By increasing the number of messages decoded at a terminal, we impose additional constraints on the achievable rate region, since the correct decoding of the interfering codewords has to be possible. The additional rate constraints might be loose in a subset of the parameter region and thus interference decoding can provide a rate advantage in some class of channels. Since (16a) and (16g) are still interference limited, it is possible that no power allocation exists to achieve a given target rate vector.

C. Partially cooperative scheme

We now consider the case in which the relay nodes cooperate in transmitting one message. This corresponds to the smallest level of cooperation which corresponds to a base station power consumption of $P^{\text{BS}} = 2 \cdot 16^{R_{\text{sym}}} - 2$.

Consider the scenario for $b > 1 > a$: since a is small it is necessary for the two RNs to cooperate

in transmitting W_2 and take advantage of the combining gain. Since $b > a$ it is more convenient to send W_3 through RN 1 and W_1 through RN 2 and to superimpose both codewords over the codeword for W_2 .

The resulting CGRAS is CGRAS number 3 in Fig. 5 (same as CGRAS number 15 Fig. 6) with achievable rate region:

$$R_1 + R_2 \leq \mathcal{C} \left(\frac{b^2 P_{21} + (b+1)^2 P_2}{1 + P_{13}} \right) \quad (19a)$$

$$R_1 \leq \mathcal{C} \left(\frac{b^2 P_{21}}{1 + P_{13}} \right) \quad (19b)$$

$$R_3 + R_2 \leq \mathcal{C} \left(\frac{b^2 P_{13} + (b+1)^2 P_2}{1 + P_{21}} \right) \quad (19c)$$

$$R_3 \leq \mathcal{C} \left(\frac{b^2 P_{13}}{1 + P_{21}} \right) \quad (19d)$$

$$R_2 \leq \mathcal{C} \left(\frac{4a^2 P_2}{1 + a^2 P_{13} + a^2 P_{21}} \right), \quad (19e)$$

with the assignment

$$X_1 = \sqrt{P_2} U_{12 \rightarrow 123}(W_2) + \sqrt{P_{13}} U_{1 \rightarrow 3}(W_3) \quad (20a)$$

$$X_2 = \sqrt{P_2} U_{12 \rightarrow 123}(W_2) + \sqrt{P_{21}} U_{2 \rightarrow 1}(W_1), \quad (20b)$$

which has a total power consumption at the access link of

$$P_1^{\text{RN}} + P_2^{\text{RN}} = 2P_2 + P_{21} + P_{13}. \quad (21)$$

Note that we have chosen the scaling of U_2 in X_1 and X_2 that provides the maximal ratio combining at RX 2. Cooperation among the RNs in transmitting W_2 provides a combining gain at all the RXs. For RX 1 and RX 3 this combining gain is $(b+1)^2 P_2$ while for RX 2 is $4a^2 P_2$, which is the largest combining gain that can be attained when fixing the total transmission power for W_2 at the RNs.

Note that, since W_2 is the base codeword, it is also decoded at RX 1 and RX 3 and thus this combining does not result in further interference for the non-intended decoders.

D. Partially cooperative scheme with interference decoding

The rate bounds in (19) are all reduced by the power of the interference experienced at each decoder. We seek to improve on this scheme by including interference decoding: we do so by allowing RX 1 and RX 3 to decode every codeword.

The resulting CGRAS number 4 in Fig. 5 (same as CGRAS number 18 in Fig. 6) which achieves the rate region:

$$R_1 + R_2 + R_3 \leq \mathcal{C} (b^2 P_{21} + (b+1)^2 P_2 + P_{13}) \quad (22a)$$

$$R_1 + R_3 \leq \mathcal{C} (b^2 P_{21} + P_{13}) \quad (22b)$$

$$R_1 + R_2 \leq \mathcal{C} (b^2 P_{21} + (b+1)^2 P_2) \quad (22c)$$

$$R_1 \leq \mathcal{C} (b^2 P_{21}) \quad (22d)$$

$$R_3 \leq \mathcal{C} (P_{13}) \quad (22e)$$

$$R_1 + R_2 + R_3 \leq \mathcal{C} (b^2 P_{13} + (b+1)^2 P_2 + P_{21}) \quad (22f)$$

$$R_1 + R_3 \leq \mathcal{C} (b^2 P_{13} + P_{21}) \quad (22g)$$

$$R_3 + R_2 \leq \mathcal{C} (b^2 P_{13} + (b+1)^2 P_2) \quad (22h)$$

$$R_3 \leq \mathcal{C} (b^2 P_{13}) \quad (22i)$$

$$R_1 \leq \mathcal{C} (P_{21}) \quad (22j)$$

$$R_2 \leq \mathcal{C} \left(\frac{4a^2 P_2}{1 + a^2 P_{13} + a^2 P_{21}} \right), \quad (22k)$$

with the assignment

$$X_1 = \sqrt{P_2} U_{12 \rightarrow 123}(W_2) + \sqrt{P_{13}} U_{1 \rightarrow 13}(W_3) \quad (23a)$$

$$X_2 = \sqrt{P_2} U_{12 \rightarrow 123}(W_2) + \sqrt{P_{21}} U_{2 \rightarrow 13}(W_1). \quad (23b)$$

As for the scheme is Sec. V-B, interference decoding improves the SINR at RX 1 and RX 3 but it does impose additional constraint on the rate of the interfering codewords being decoded.

In particular, for $b > 1 > a$, the rate of R_1 (R_3) is reduced to $\mathcal{C}(P_{21})$ ($\mathcal{C}(P_{13})$) from $\mathcal{C}(b^2 P_{21})$ ($\mathcal{C}(b^2 P_{13})$) since correct decoding must take place at RN 3 (RN 1) as well.

VI. SIMULATION SETUP

Through numerical simulation, we can consider a large number of schemes and select the scheme corresponding to the smallest power consumption. The number of achievable transmission strategies which can be generated using [1, Th. IV.2.] is on the order of hundreds without rate-splitting and thousands with rate-splitting. The power optimization for each scheme can be efficiently performed using linear optimization tools under some mild restrictions on the power allocation for the cooperative schemes.

Indeed the number of schemes considered and the low computational complexity make it possible to identify a very detailed transmission strategy with reasonable resources.

In the following we consider the following numerical optimization problems:

- **No rate-splitting Scenario:** When no rate-splitting is considered, the number of possible achievable rate regions is limited and it is possible to graphically present the results of the energy efficiency optimization for varying a and b parameters. We also connect the numerical results with the schemes in Sec. V to provide further intuitions.
- **Rate-Splitting Scenario:** when considering both superposition coding and rate-splitting there are simply too many schemes to be efficiently presented graphically. In this case we group the schemes according to their features in terms of the cognition level, number of superposition coding steps and amount of interference decoding. This description is less detailed than the previous scenario, but still insightful on a higher level perspective as in Sec. III.
- **Comparison between Inner and Outer Bounds:** we compare the results of the numerical optimization with the lower bound on the energy consumption in Sec. IV and the case of relay selection and uncoordinated transmissions. For this scenario we consider the throughput comparison among different strategies for fixed channel parameters.

A. No Rate-Splitting Scenario

We begin by considering the achievable schemes that can be generated from [1, Th. IV.2.] which do not involve rate-splitting. By restricting our attention to this class of schemes, the optimization results are limited to eighteen schemes (after accounting for symmetry). Given the limited set of schemes, we can provide various intuitions on the important features of the optimal solution for varying a and b .

The solution of the optimal energy efficiency problem are represented using the corresponding CGRAS as in Fig. 6. Each CGRAS is contained in a different partition of the figure with an associated number and color: The schemes are arranged by increasing level cooperation and increasing interference decoding:

- **blue tones, numbers 1–8:** no cooperation and minimum or no interference decoding schemes,
- **green and yellow tones, numbers 9–13:** no cooperation and interference decoding schemes,
- **red tones, numbers 14–18:** partial cooperation schemes with some interference decoding.

In each box, two rectangles represent the two RNs: the rectangle with a red, dotted edge is associated with RN 1 while the one with a blue, solid edge to RN 2. Each rectangle contains the CGRAS nodes

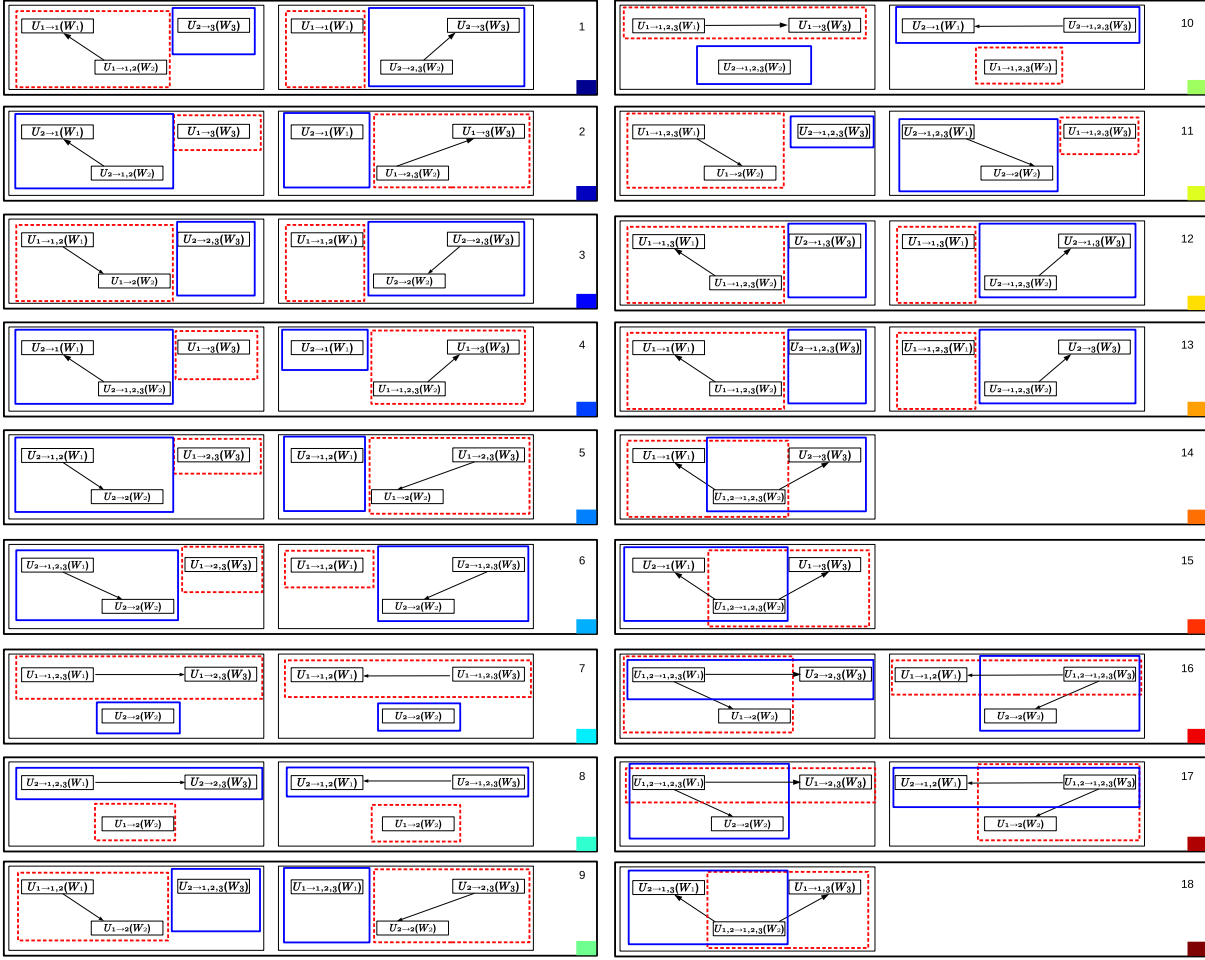


Fig. 6. A graphical representation of the best performing CGRASs for the scenario in Sec. VI-A.

known at the associated RN. If a codeword is known at both, it is placed in the intersection of the two rectangles. Each node of the CGRAS is represented by a box labeled as $U_{i \rightarrow j}(W_z)$. As in [1, Def. 1], $U_{i \rightarrow j}$ is the RV associated with the codeword transmitted from the set of transmitter i toward of the set of receiver j while W_z indicates that this codeword embeds the message W_z as in Sec. V. Superposition coding is indicated as a directed arrow, from the base codewords to the top one. Due to the symmetry in the channel model and in the target rate vector, more than one scheme achieves the same achievable rate region. In particular, given any scheme, swapping the labeling of RN 1 and RN 3 and swapping 2 to 1 in i and 1 and 3 in j for all the codewords $U_{i \rightarrow j}$ will result in a scheme with the same achievable rate region. We refer to these schemes as equivalent and equivalent schemes are placed in the same partition of the figure.

The optimization is performed as follows: for each possible message allocation among the RNs, all possible combinations of superposition coding and interference decoding are considered. With each such choice, a CGRAS is obtained along with the corresponding achievable rate region through [1, Th. IV.2.].

The achievable rates are evaluated for the choice of auxiliary RV in [1, Lem. IV.3.], additionally the scaling of the cooperative codewords is chosen so as to provide the largest combining gain at the intended receiver. With this choice of distribution of $U_{i \rightarrow j}$, the power consumption is a linear function of the matrix A in [1, eq. (27)] and therefore the minimum power consumption on the access link for which the target symmetric rate is achievable can be determined using standard linear programming algorithms. The total power consumption is obtained by adding the power consumption at the access link with the power consumption on the relay link which is obtained through [1, Th. V.1.] for each message allocation at the RNs. This optimization is repeated over multiple channel parameters (a, b) . For each channel, the optimal schemes are selected with a 5% tolerance and when more than one scheme is within the prescribed tolerance, we choose the scheme which is optimal for values of similar of (a, b) . In the following, we consider the four symmetric rates: $R_{\text{sym}} \in [0.1, 0.5, 1, 2]$ and the range of channel parameters $a \times b = [0 \dots 2]^2$.

Let's start by considering the case of $R_{\text{sym}} = 0.1$: the optimization results are provided in Fig. 7(a). Since the achievable rate is relatively low, we expect the scheme of Sec. V-A to be both feasible and efficient in achieving R_{sym} . Indeed scheme 1 and 2, which correspond to the encoding choice of Sec. V-A, are optimal around the region $b > a > 1$ and the region $b < 1 = a$. In these two regions the level of the interference at each receiver is limited and treating the interference as noise is optimal. Also, there is no need for cooperation as the advantages provided by the combining gains do not outweigh the additional costs of cooperation on the relay link. A detailed explanation of scheme 1 is provided in Sec. V-A. For the region $b > a > 1$, we notice that RN 1/ (RN 2) can communicate to RX 3/ (RX 1) through a link with gain b at high SNR. RX 2 can be served by either RN 1 or RN 2 also through the link a : the presence of simultaneous transmission on this link creates excessive interference to the rate R_2 since a large b guarantees that the power of this transmission is low.

For larger values of the parameter a , it is more advantageous for RX 2 to decode part of the interference and thus schemes similar to the one in Sec. V-B perform better. Indeed schemes 3 and 5, having the same coding strategy of the scheme in Sec. V-B, are optimal around the region $a > 1 > b$ and $a > b > 1$. The scheme 3 is presented in Sec. V-B, scheme 5 is a simple variation of scheme 3 where RN 1 (RN 2) serves RX 3 (RX 1).

When a is small, then coherent combining is necessary to attain $R_2 = R_{\text{sym}}$ and the RNs need to cooperate in transmitting message RX 2. The remaining two messages are superimposed over the codeword for W_2 and we thus obtain the cooperative scheme of Sec. V-C and of scheme 14 and scheme 15. Note

that the value of b only determines which RN serves which RX and no interference decoding is required at RX 1 and RX 2 apart for the case $b \approx 1$.

Let's consider now the case $R_{\text{sym}} = 0.5$ presented in Fig. 7(b). Since the symmetric rate is higher than the case depicted in Fig. 7(a), it is no longer optimal to treat the interference as noise and thus the region in which scheme 1 and scheme 2 are optimal are smaller than in Fig. 7(a). Schemes involving various possible interference decoding, such as scheme 4, 7, 8 and 12 emerge as optimal. Also Scheme 18, which is detailed in Sec. V-D and which involves cooperation and interference decoding, becomes optimal in a larger region around $b \approx 1$.

As the rate R_{sym} is further increased to $R_{\text{sym}} = 1$ and $R_{\text{sym}} = 2$ in Fig. 8(a) and 8(b) we see that cooperation becomes more and more advantageous as it allows the RNs to apply superposition coding in order to manage the interference from simultaneous transmissions. When applying superposition coding, the receiver of the top codeword also decodes the bottom codewords. In doing so, this receiver also performs interference decoding and can thus partially remove the effect of the interference. Although costly on the access link, cooperation makes it possible to perform two superposition coding steps instead of one, and this is the key of the efficiency of schemes 14, 16, 17 and 18.

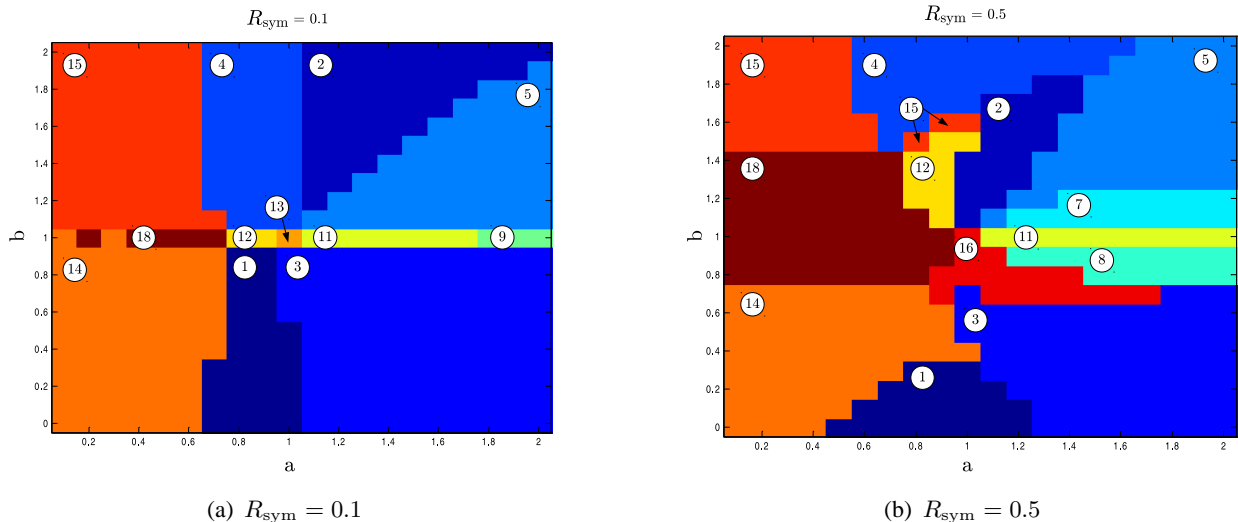


Fig. 7. Numerical simulations for the best CGRAS performance for $R_{\text{sym}} = 0.1$ and $R_{\text{sym}} = 0.5$ for $a \times b \in [0, 2]^2$.

Fig. 9 plots the overall power consumption for the optimization results in Fig. 7(a), 7(b), 8(a) and 8(b). In all cases the largest power consumption is attained at small values of a : since both RNs have a link of gain a toward RX 2, it is not possible to obtain a better SNR toward this node. Since the communication toward RX 2 is arduous for both receiver in this regime, cooperation is important to

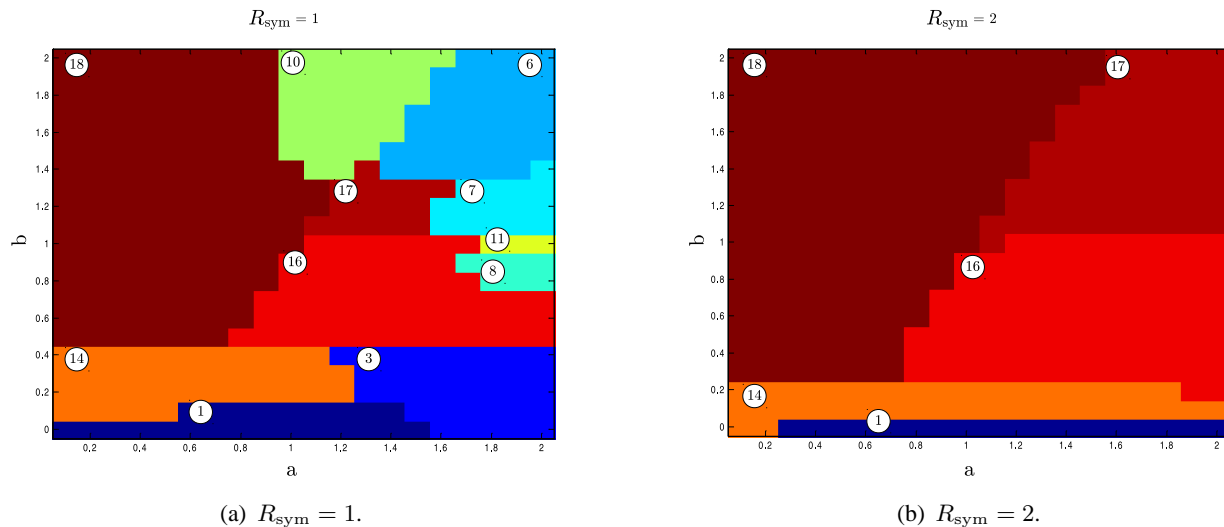


Fig. 8. Numerical simulations for the best CGRAS performance for $R_{\text{sym}} = 1$ and $R_{\text{sym}} = 2$ for $a \times b \in [0, 2]^2$.

provide combining gains which indeed greatly reduces the power consumption from moderate values of a . Another regime with high power consumption for large symmetric rates is for small a and around $0.2 < b < 0.8$: by referring to the optimization results we see that this is a transition point between scheme 14 and scheme 18. In both schemes the two relays cooperate in transmitting the codeword for W_2 and superimpose the messages for RX 1 and RX 3 over this codeword. The two schemes differ in that scheme 14 does not employ interference decoding at RX 1 and RX 3 while scheme 18 does. Since b is relatively small, the power at which the codewords for RX 3 is received at RX 1 is not sufficiently strong to be decoded and yet too strong to be treated as noise. The same occurs at RX 3 for the interfering codeword for RX 1. The solution to this impasse is for both decoders to increase their transmitted power, so as to eventually increase the received SNR. From a high level perspective, and given the symmetry in the channel and the target rate, the SNR at both RX 1 and RX 3 scales as $P/(1 + b^2 P)$ when decoding W_2 but treating the remaining interference as noise. Clearly an increase in P increases the SNR, although very slowly. For this reason the power consumption in this region is relatively high: a high power P is required to attain the SNR which grants reliable communication at rate R_{sym} .

The power consumption corresponding to the numerical optimization in Fig. 7(a), 7(b), 8(a) and 8(b) are shown in Fig. 9(a), 9(b), 9(c) and 9(d) respectively. From these figures is clear that, as the symmetric rate increases, the largest power consumption corresponds to the region $a \in [0, 1]$ and $b \in [0.2, 1]$: which corresponds to the region where the numerical optimal scheme changes from 14 to 18. In scheme 14 the codeword for W_3 (W_1) is treated as noise RX 1 (RX 3) while in scheme 18 this codeword is decoded.

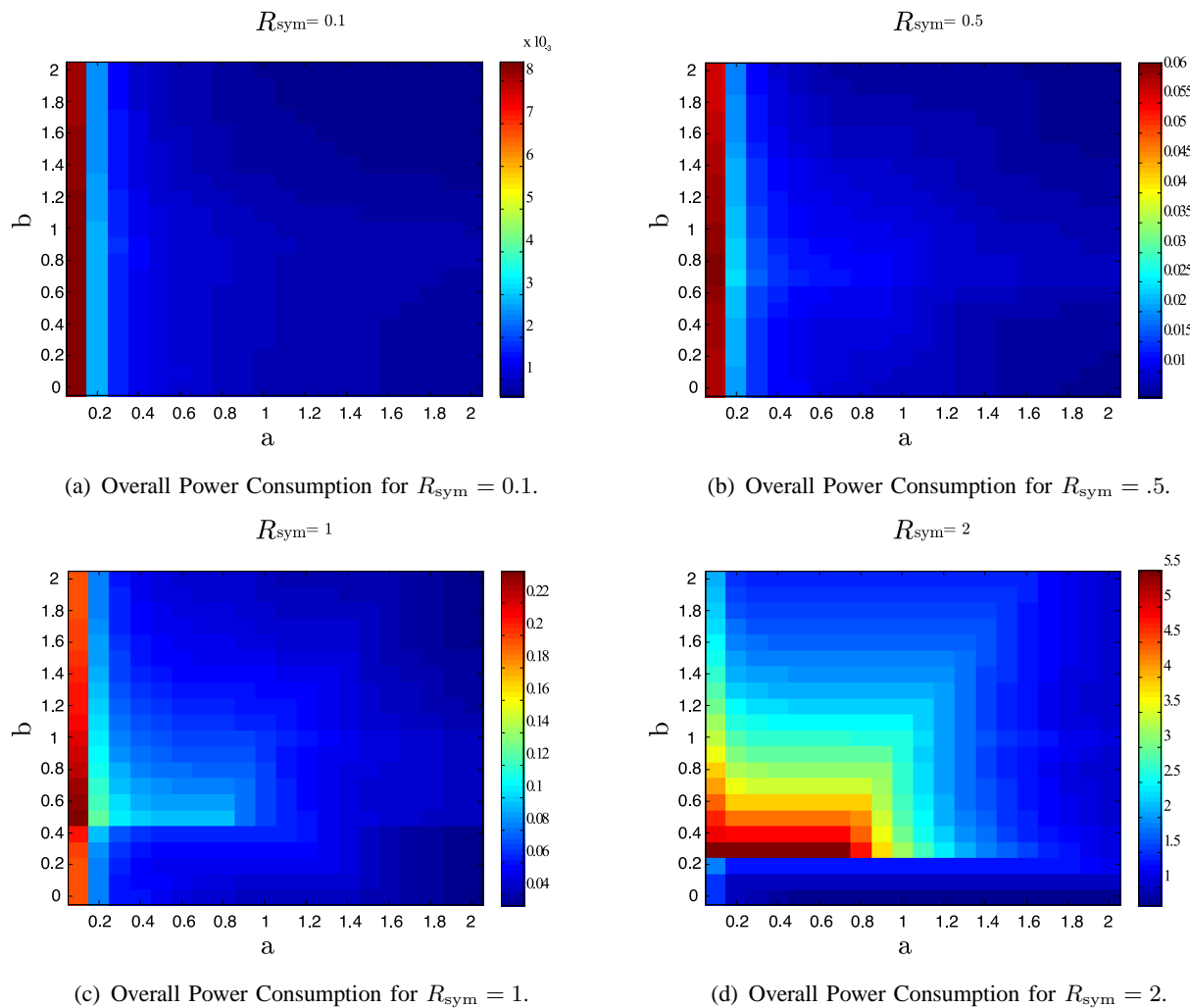


Fig. 9. Power efficiency for varying symmetric rate R_{sym} and $a \times b \in [0, 2]^2$.

In this region, therefore, the large power consumption corresponds to the tension between treating the interference as noise or decoding the interference which increases as the symmetric rate increases. As we shall see, rate-splitting makes it possible to reduce the power consumption in this region by allowing partial interference cancellation.

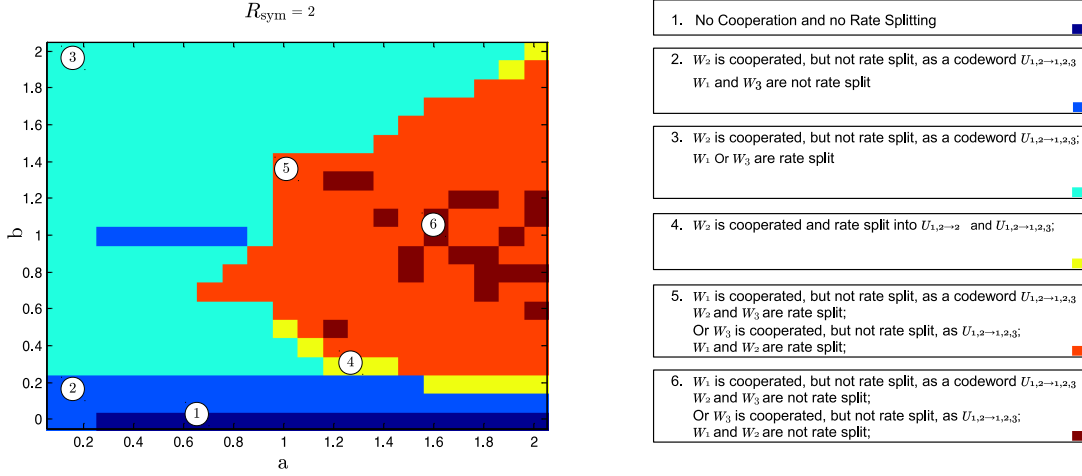
B. Rate-Splitting Scenario

In this section we analyze the effect of rate-splitting on the power consumption. Since the number of schemes that can be generated using [1, Th. IV.2.] in this case is quite high, it is not longer possible to analyze each scheme separately as in Sec. VI-A. Instead we focus on higher-level features of the transmission schemes and overall power consumption.

The solution of the power minimization with rate-splitting for the case $R_{\text{sym}} = 2$ is presented in Fig. 10(a) while the legend of this figure is in Fig. 10(b). The optimal schemes are numbered and color coded

according to the level of cooperation among the RNs, in particular:

- **blue tones, numbers 1 and 2:** no cooperation with either with or without rate-splitting,
- **red tones, numbers 3–5:** partial cooperation schemes with some interference decoding.
- **yellow tones, numbers 6 and 7:** no cooperation and interference decoding schemes.



(a) Numerical simulations for the best CGRAS performance considering rate-splitting for $R_{\text{sym}} = 2$ and $a \times b \in (0, 2)^2$.

(b) CGRAS Legend for rate-splitting scenario.

Fig. 10. Best CGRAS performance with rate splitting.

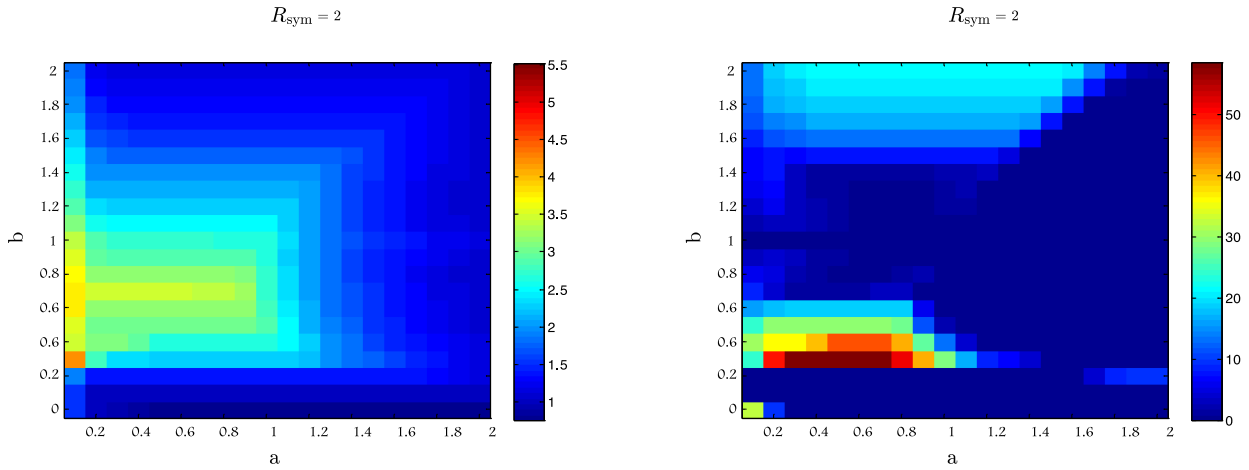
From Fig. 10(a) it is clear that rate-splitting always provide an advantage: the schemes which do not employ rate-splitting, scheme 1 and scheme 2, are optimal mostly in the regime where b is close to zero, which corresponds to the case where no interference is created at RX 1 and RX 2. As b increases, rate-splitting allows for the receiver to decode part of the interfering codeword and strip it from the received signal. For value of $a < 1$, the optimal scheme often involve cooperation over W_2 : since the link toward RX 2 is relatively weak, the RNs need not cooperate in the transmission of the codeword associated to this message. As the power use to transmit this codeword combines coherently at all the RXs, it is more advantageous for all RXs to decode it, as in scheme 14 and 18 in Fig. 8(b), and no rate-splitting is necessary. rate-splitting is instead useful in the transmission of W_1 and W_3 as it allows a RX to decode part of the interfering codeword.

For the case $a > 1$, cooperation is no longer necessary to transmit W_2 and is instead useful when sending W_1 or W_3 : as for scheme 16 and 17 in Fig. 8(b) the codeword for either message is decoded at all the receivers and acts as a base codeword to transmit the other messages while, again, rate-splitting is useful in allowing for partial cancelation of interfering codewords.

The power consumption associated with the numerical optimization in Fig. 8(b) is presented in Fig.

11(a) and can be compared to the non rate-splitting power consumption in Fig. 9(d). As for Fig. 9(d), the largest power consumption is in the range $a \in [0, 1]$ and $b \in [0.2, 1]$. The difference is the two figures is the regime $a \approx 0.3$: without rate-splitting, this regime has the largest power consumption, whether with rate-splitting the power consumption in this regime is greatly reduced.

Fig. 11(b) plots the power efficiency improvement of rate-splitting case with respect to the non rate-splitting case. The power improvement are relatively small for the class of channel in which a low power consumption can already be attained without rate-splitting. In the cases in which the performance of the system are more severely limited by the interference, rate-splitting attains substantial power efficiency gains.



(a) Power efficiency with rate-splitting for $R_{\text{sym}} = 2$ and $a \times b \in (0, 2)^2$. (b) Performance comparison of rate-splitting and no rate-splitting scenarios for $R_{\text{sym}} = 2$ and $a \times b \in (0, 2)^2$.

Fig. 11. Power efficiency with rate-splitting

C. Comparison between Inner and Outer Bounds

In this section we compare the lower bound in Sec. IV with the upper bound in [1, Sec. IV.B.]. The upper bound is evaluated through numerical optimization while the lower bound is evaluated through Monte-Carlo simulations. A plot of the upper and lower bound on the energy consumption for the channel model in Fig. 4 for $a = b = 1$ and increasing symmetric rate R_{sym} is presented in Fig. 12. This figure shows that transmission schemes involving rate-splitting provides clear energy saving over the schemes without rate-splitting, especially at high symmetric rates. Unfortunately, the gap between the best upper bound and the lower bound increases as the symmetric rate increases: the lower bound in Sec. IV is known to be loose for the general channel model, so the divergence between upper and lower bounds is expected.

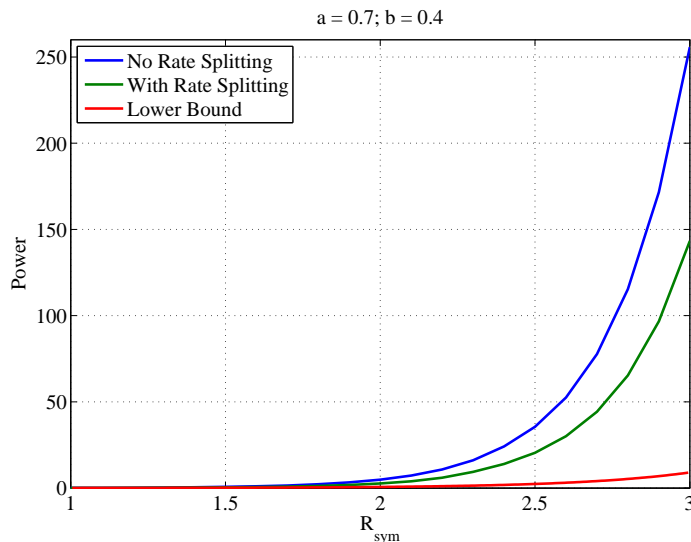


Fig. 12. Performance comparison of rate-splitting and no rate-splitting scenarios with outer bound for $R_{\text{sym}} \in [1, 3]$, $a = 0.7$ and $b = 0.4$.

D. Relevance of the Results for Practical Networks

We conclude the section by drawing insight from the numerical simulation on practical communication scenarios. The network in Fig. 4 well models a practical, LTE-style cellular networks in which RX 2 is a receiver located at the coverage edge among relay nodes while RX 1 and RX 3 are users served in a good network coverage area. Since RX 2 is at the edge of two cells, a reasonable scenarios would be to assume a small. Also, since RX 1 and RX 3 are within the cell boundary, we expect b to be small.

End users operating at the edge of coverage area of two or more nodes are a common and problematic scenario in cellular networks, especially in LTE-A networks. While in GSM networks this issue was addressed by frequency reuse planning, in the predecessor of LTE-A, UMTS networks, such cases were dealt by so-called soft handovers, where the cell edge users were served by more than one cell simultaneously and selection combining was compensating for poor coverage and interference, decreasing the power consumption at each cell as well as the interference toward the users with good network coverage [22]. Due to architectural changes and faster scheduling demands, which limits the cooperation among the base stations, soft handover is no more an available feature in LTE [23], leaving the issue of users on the coverage borders an open problem. The research needed to address this problem is ongoing for future LTE releases and several possible solutions were described in this work. Solutions like fractional frequency reuse [24], [25] are being considered but are not efficient as they limit the frequency spectrum allocation and, consequently, user throughputs (although making the frequency allocation more flexible).

Cooperative strategies, such as CoMP, represent the most promising way of optimizing the performance

of coverage edge users. Cooperation among serving nodes is becoming feasible for the next generation of wireless networks. The new decentralized network architecture of LTE poses difficulties in cooperation taking place among multiple base stations, due to delay issues. On the other hand, the cooperation among the nodes served by a single base station is a more feasible and promising solution [26].

Considering these trends, the scenario under our consideration is indeed very relevant for practical applications.

From the numerical simulation of the CGRAS without rate-splitting in Sec. VI-A, we conclude that cooperation among RNs toward the cell edge user RX 2 is optimal. In fact, regardless of the symmetric rate, in the region for small a and small b the optimal schemes are 14, 15 and 18, all of which consider cooperation over W_2 .

The drawback of cooperation is the tradeoff between relay link and access link power consumption and the resource allocation. In particular, spectrum availability at the relay link may become a bottleneck in practical systems, assuming the base station also serves some end users directly. However, the common target in network planning is to minimize the coverage edge areas and for the coverage overlap among two and three cells not to exceed 30% and 10% respectively of the total covered area, making the cooperation for edge users less resource consuming from the relay link point of view.

Moreover, the simulation results show it is optimal set the codeword of the edge user as a bottom codeword for superposition, which implies that it is optimal for the users in a better coverage of both relay nodes to decode the message transmitted to the edge user.

Similar conclusions can be drawn from the simulation with rate-splitting: key areas where cooperation is optimal are the areas with lower parameter a values in which again the codeword for RX 2 acts as a bottom codeword for superposition and is therefore decoded by the other receivers.

VII. CONCLUSION

The relationship between cooperation and energy efficiency in relay-assisted downlink cellular system is studied through an information theoretical approach. We consider in particular the case in which a base station communicates to the three receivers through the aid of two relays. We allow for the base station to transmit the message of one receiver to multiple relay nodes which makes it possible for the relay nodes to partially coordinate their transmission. This scenario idealizes LTE-style cellular networks in which relay nodes are used to improve the energy efficiency of the network. We present upper and lower bounds to the overall energy consumption in this network which provide important insight on energy

efficient transmission schemes for larger networks. Our results clearly show that relay cooperation is necessary in attaining a higher power efficiency along with high overall network throughput. In particular cooperation is necessary when transmitting toward users on the cell edge and which have low gain toward multiple relays. We also present a new numerical simulation tool which generates a large number of achievable rate regions based on superposition coding, rate splitting and interference decoding. With this tool we can consider a large number of transmission strategies and choose the most power efficient. The power optimization of each achievable rate region can be performed in with low complexity and thus this approach can be easily extended to larger networks.

ACKNOWLEDGMENTS

The authors would like to thank Prof. Gerhard Kramer for the stimulating conversations and useful comments.

REFERENCES

- [1] S. Rini, E. Kurniawan, L. Ghaghanidze, and A. Goldsmith, "Energy efficient cooperative strategies for relay-assisted downlink cellular systems, part I: Theoretical framework," *IEEE Journal on Selected Areas in Communications (JSAC)* - submitted, 2013.
- [2] R. Pabst, B. Walke, D. Schultz, P. Herhold, H. Yanikomeroglu, S. Mukherjee, H. Viswanathan, M. Lott, W. Zirwas, M. Dohler *et al.*, "Relay-based deployment concepts for wireless and mobile broadband radio," *Communications Magazine, IEEE*, vol. 42, no. 9, pp. 80–89, 2004.
- [3] J. Luo, R. Blum, L. Cimini, L. Greenstein, and A. Haimovich, "Link-failure probabilities for practical cooperative relay networks," in *Vehicular Technology Conference. 'VTC'. IEEE 61st*, vol. 3, may - june 2005, pp. 1489 – 1493 Vol. 3.
- [4] Z. Lin, E. Erkip, and A. Stefanov, "Cooperative regions for coded cooperative systems [mobile radio systems]," in *Global Telecommunications Conference. 'GLOBECOM'. IEEE*, vol. 1, nov.-3 dec. 2004, pp. 21 – 25 Vol.1.
- [5] Z. Li and E. Erkip, "Relay search algorithms for coded cooperative systems," in *Global Telecommunications Conference. 'GLOBECOM'. IEEE*, vol. 3, nov.-2 dec. 2005, p. 6 pp.
- [6] Y. Zhao, R. Adve, and T. Lim, "Improving amplify-and-forward relay networks: optimal power allocation versus selection," in *Information Theory, 2006 IEEE International Symposium on*. IEEE, 2006, pp. 1234–1238.
- [7] N. Devroye, P. Mitran, and V. Tarokh, "Cognitive multiple access networks," in *Proc. IEEE International Symposium on Information Theory (ISIT), Adelaide, Australia, 2005*, pp. 57–61.
- [8] O. Sahin and E. Erkip, "Achievable rates for the Gaussian interference relay channel," in *Proc. IEEE Global Telecommun. Conf.*, 2007, pp. 1627–1631.
- [9] A. Bletsas, A. Khisti, D. Reed, and A. Lippman, "A simple cooperative diversity method based on network path selection," *Selected Areas in Communications, IEEE Journal on*, vol. 24, no. 3, pp. 659 – 672, march 2006.
- [10] A. Stefanov and E. Erkip, "Cooperative space-time coding for wireless networks," *Communications, IEEE Transactions on*, vol. 53, no. 11, pp. 1804–1809, 2005.

- [11] N. Lee, H. Yang, and J. Chun, "Achievable sum-rate maximizing af relay beamforming scheme in two-way relay channels," in *Communications Workshops. ICC Workshops. IEEE International Conference on*. IEEE, 2008, pp. 300–305.
- [12] B. Rankov and A. Wittneben, "Achievable rate regions for the two-way relay channel," in *Information Theory, IEEE International Symposium on*. IEEE, 2006, pp. 1668–1672.
- [13] W. Chen, Y. Yuan, C. Xu, K. Liu, and Z. Yang, "Virtual mimo protocol based on clustering for wireless sensor network," in *Computers and Communications, 2005. ISCC 2005. Proceedings. 10th IEEE Symposium on*, june 2005, pp. 335 – 340.
- [14] A. Coso, S. Savazzi, U. Spagnolini, and C. Ibars, "Virtual mimo channels in cooperative multi-hop wireless sensor networks," in *Information Sciences and Systems, 2006 40th Annual Conference on*, march 2006, pp. 75 –80.
- [15] R. Fantini, D. Sabella, and M. Caretti, "An e3f based assessment of energy efficiency of relay nodes in lte-advanced networks," in *Personal Indoor and Mobile Radio Communications (PIMRC), 2011 IEEE 22nd International Symposium on*, sept. 2011, pp. 182 –186.
- [16] R. Fantini, M. Caretti, and D. Sabella, "Energy efficiency in lte-advanced networks with relay nodes," *Vehicular Technology Conference ('VTC' Spring), 2011 IEEE 73rd*, 2011.
- [17] Y. Xiao and L. Cimini, "Energy efficiency of distributed cooperative relaying," in *MILITARY COMMUNICATIONS CONFERENCE, 2011-MILCOM 2011*. IEEE, 2011, pp. 73–78.
- [18] R. Madan, N. Mehta, A. Molisch, and J. Zhang, "Energy-efficient cooperative relaying over fading channels with simple relay selection," *Wireless Communications, IEEE Transactions on*, vol. 7, no. 8, pp. 3013–3025, 2008.
- [19] O. Sahin and E. Erkip, "Achievable rates for the Gaussian interference relay channel," in *Proc. IEEE Global Telecommun. Conf.*, 2007, pp. 1627–1631.
- [20] S. Rini, D. Tuninetti, and N. Devroye, "Inner and outer bounds for the Gaussian cognitive interference channel and new capacity results," *Information Theory, IEEE Transactions on*, vol. 58, no. 2, pp. 820–848, 2012.
- [21] —, "New results on the capacity of the gaussian cognitive interference channel," in *Communication, Control, and Computing (Allerton), 2010 48th Annual Allerton Conference on*. IEEE, 2010, pp. 637–644.
- [22] M. Schinnenburg, I. Forkel, and B. Haverkamp, "Realization and optimization of soft and softer handover in umts networks," in *Personal Mobile Communications Conference, 2003. 5th European (Conf. Publ. No. 492)*. IET, 2003, pp. 603–607.
- [23] A. Racz, A. Temesvary, and N. Reider, "Handover performance in 3gpp long term evolution (lte) systems," in *Mobile and Wireless Communications Summit, 2007. 16th IST*. IEEE, 2007, pp. 1–5.
- [24] A. Stolyar and H. Viswanathan, "Self-organizing dynamic fractional frequency reuse in ofdma systems," in *INFOCOM 2008. The 27th Conference on Computer Communications*. IEEE. IEEE, 2008, pp. 691–699.
- [25] H. Lei, L. Zhang, X. Zhang, and D. Yang, "A novel multi-cell ofdma system structure using fractional frequency reuse," in *Personal, Indoor and Mobile Radio Communications, 2007. PIMRC 2007. IEEE 18th International Symposium on*. IEEE, 2007, pp. 1–5.
- [26] M. Sawahashi, Y. Kishiyama, A. Morimoto, D. Nishikawa, and M. Tanno, "Coordinated multipoint transmission/reception techniques for lte-advanced [coordinated and distributed MIMO]," *Wireless Communications, IEEE*, vol. 17, no. 3, pp. 26–34, 2010.